

El uso de software libre en el desarrollo del proyecto “UPCommons: portal de acceso abierto al conocimiento de la UPC”

Jordi Prats, Toni Prieto, Oriol Rico

Jordi.Prats@upc.edu, Antonio.Juan.Prieto@upc.edu, Oriol.Rico@upc.edu

Universitat Politècnica de Catalunya (UPC)
Servei de Biblioteques i Documentació
Jordi Girona 31
08034 Barcelona

Abstract:

El Servicio de Bibliotecas y Documentación de la Universidad Politécnica de Cataluña (UPC) ha ido desarrollando distintos repositorios institucionales para facilitar la publicación, en acceso abierto, de los trabajos realizados por la comunidad universitaria. En el 2006 todos los repositorios confluyen en un proyecto común: *UPCommons*: portal de acceso abierto al conocimiento de la UPC.

Para el desarrollo del proyecto ha sido de gran importancia poder disponer de herramientas tanto de programación como paquetes ya desarrollados en código abierto y licencias GPL, entre las cuales destaca el programa DSpace, así como de estándares abiertos que han facilitado la implementación de modelos de interoperabilidad con otros sistemas de información de la universidad.

Palabras clave:

Repositorios institucionales, acceso abierto, DSpace, software libre, institutional repositories, open access, open source

1.- Introducción

La Universidad Politécnica de Cataluña (UPC) empezó su andadura en el desarrollo de repositorios institucionales en el 2001 participando, con el Consorcio de Bibliotecas Universitarias de Cataluña (CBUC), en la puesta en funcionamiento del portal TDXⁱ (“Tesis Doctorals en Xarxa”), un proyecto cooperativo de todas las universidades públicas catalanas que tenía, y tiene aún, el objetivo de publicar en acceso abierto las tesis leídas en los diferentes centros.

Tomando como punto de partida este proyecto y siguiendo el auge que el movimiento Open Access ha tenido en los últimos años, en el 2002 se empiezan a desarrollar en la propia universidad repositorios que tienen por objetivo facilitar a los autores la publicación de sus trabajos académicos en acceso abierto para potenciar su visibilidad, a la vez que se pretende garantizar su preservación.

El desarrollo de distintos repositorios, que se presentan más adelante, orientados a la gestión de recursos específicos conllevó una cierta dispersión de sus contenidos y se detectó la necesidad de crear un portal que permitiese el acceso a todos los trabajos de manera unificada, a la vez que se gestaba un modelo que facilitara la identificación de las distintas iniciativas en un único concepto y proyecto. Fue en este marco que se creó el proyecto y portal **UPCommons**: portal de acceso abierto al conocimiento de la UPCⁱⁱ.

Para el desarrollo de las distintas iniciativas englobadas dentro de **UPCommons** ha sido de gran importancia el poder contar con herramientas de código abierto, así como de estándares también abiertos y ampliamente aceptados. Ello ha facilitado y acelerado el desarrollo de proyectos que, si bien tienen una carga tecnológica importante, dependen en gran medida de aspectos organizativos en instituciones grandes como pueden ser las universidades. Así mismo los estándares abiertos han facilitado a su vez su integración con otros sistemas de información de la universidad, aspecto importante para la consolidación de algunos de los repositorios.

2.- **UPCommons**: portal de acceso abierto al conocimiento de la UPC

Página principal de
UPCommons: portal de acceso abierto al conocimiento de la UPC

UPCommons: portal de acceso abierto al conocimiento de la UPC es el proyecto que engloba el desarrollo de repositorios institucionales abiertos de la UPC, a la vez que da nombre al portal que facilita el acceso de manera unificada a todos sus contenidos e informa a la comunidad universitaria sobre los servicios de soporte que encontrará en las bibliotecas que le van a facilitar la publicación, en acceso abierto, de sus trabajos académicos.

Los objetivos del proyecto no difieren mucho de los de otras iniciativas similares, focalizándose principalmente en incrementar la visibilidad de la actividad de investigación y docencia de la universidad en la red; intentar garantizar la preservación de los trabajos resultantes de esta actividad, e informar y asesorar sobre los aspectos relacionados con la propiedad intelectual que lleva la publicación en acceso abierto de estos trabajos.

En este marco se han desarrollado seis repositorios orientados cada uno de ellos a alojar distintas tipologías de documentos, según su naturaleza en la actividad de la universidad o sus características técnicas.

A continuación se va a realizar una pequeña descripción de cada uno de ellos:

E-prints UPCⁱⁱⁱ

Tiene por objetivo facilitar un espacio al personal investigador de la universidad para que publique en acceso abierto cualquier trabajo resultante de su investigación listo para entrar en un proceso de comunicación.

Se estructura en función de las áreas de conocimiento propias de la UPC, bajo las cuales se generan subcomunidades correspondientes a unidades básicas o grupos de investigación de la universidad.

Todos los documentos se encuentran bajo alguna de las modalidades de licencias Creative Commons^{iv}. Los propios autores escogen cuál de ellas prefieren.

Revistas UPC^v

Nace en su origen como portal de acceso a los artículos de las revistas que publican las unidades básicas de la universidad, pero la demanda de publicación de los trabajos presentados en congresos organizados en la UPC hace evolucionar el producto hacia un portal de acceso a contenidos en los que, si bien los autores no son necesariamente miembros de la UPC, sí que esta actúa como entidad editora.

En este marco, el repositorio no sólo tiene por objetivo alojar trabajos ya publicados, sino también servir también de base para la creación de nuevas publicaciones. Cabe destacar que en el caso de publicaciones ya existentes no se pretende en ningún momento sustituir sus correspondientes portales, sino facilitar la preservación de sus contenidos.

TDX^{vi}

Como ya se ha dicho TDX es el primer repositorio de acceso abierto en el que participó la UPC. Con el ánimo de crear un portal de acceso abierto para las tesis doctorales leídas en las universidades catalanas, TDX nace en el 2001 siguiendo los criterios marcados por proyectos internacionales como la Networked Digital Library of Theses and Dissertations^{vii}.

En la actualidad se está facilitando la incorporación de tesis doctorales de otras universidades españolas, con lo que TDX (Tesis Doctorales en Xarxa) ha evolucionado hacia TDR (Tesis Doctorales en Red) ampliando de este modo su horizonte.

Trabajos académicos^{viii}

En el 2005 se puso en funcionamiento el repositorio de Proyectos de Fin de Carrera, Trabajos de Fin de Carrera y Tesinas. En sus inicios el proyecto se centró en la implementación de una herramienta propia desarrollada con VBScript sobre una base de datos Microsoft SQL Server.

Recientemente se ha realizado la migración de todos sus contenidos a una plataforma DSpace, hecho que se ha aprovechado para normalizar algunos de los procedimientos diseñados para su gestión y ampliar los contenidos del repositorio a otros trabajos académicos realizados por estudiantes de la universidad.

Archivo Gráfico de la Escuela Técnica Superior de Arquitectura de Barcelona^{ix}

Se trata del primer repositorio institucional desarrollado en la UPC. El Archivo Gráfico de la ETSAB conserva una parte importante de la documentación generada por la actividad académica de la escuela. En el año 2002 se inició un proyecto de digitalización de todos sus fondos y se desarrolló una herramienta para facilitar su gestión y consulta en Internet.

Videoteca Digital de la UPC^x

Se trata de un proyecto de colaboración con el Instituto de Ciencias de la Educación (ICE) con el objetivo de contar con un espacio donde alojar los vídeos digitales que se realizan en el ámbito de los proyectos de innovación docente.

Actualmente una parte importante de sus contenidos consta de grabaciones de los actos académicos que realizan los distintos centros de la universidad, así como de los vídeos que el Servicio de Comunicación y Promoción produce para difundir la propia UPC o algunos de sus programas.

En el ámbito de la innovación docente cabe destacar el papel que están realizando las Factorías de Recursos Docentes^{xi}, ya que son muchas las grabaciones que se producen en estos espacios disponibles en todas las bibliotecas de la UPC.

Proyectos en curso

En la actualidad existen dos proyectos en curso que servirán para complementar los repositorios ya existentes y facilitar a los autores de la universidad espacios para publicar, prácticamente, cualquier trabajo que estén realizando.

Por un lado se está trabajando en la implementación de un repositorio de imágenes, muy orientado a la docencia, para alojar principalmente colecciones de fotografías o diapositivas realizadas por el PDI o PAS de la universidad y que son de uso corriente en su actividad docente.

Por otro, también está en fase de desarrollo el UPC OpenCourseWare que, siguiendo el modelo marcado por el MIT^{xii}, tiene por objetivo facilitar la publicación en acceso abierto de los materiales docentes producidos en la universidad en su diversidad de estudios.

3.- Software libre

3.1.- DSpace

En el año 2004, en el marco del Consorcio de Bibliotecas Universitarias de Cataluña (CBUC)^{xiii}, se creó un grupo de trabajo en el que participaron un total de seis universidades miembros del consorcio, que tenía por objetivo la selección de un programa de software libre que sirviera tanto para el desarrollo de repositorios colectivos en el propio CBUC como para los repositorios institucionales que pudieran desarrollar las distintas universidades^{xiv}.

Se analizaron un total de siete aplicaciones (ARNO, CDSware, DSpace, EPrints, Fedora, i-Tor y MyCoRE), las cuales fueron instaladas y testeadas por los miembros del Grupo, para seleccionar finalmente el software DSpace^{xv}. Los motivos principales por los cuales se escogió esta aplicación fueron principalmente que se trataba de un producto muy acabado, con unos costes de implementación razonables e instalado en un número destacable de organizaciones importantes, hecho que garantiza su continuidad.

DSpace sale a la luz en el 2002 y ha sido desarrollado conjuntamente entre el MIT y HP. Se trata del resultado del trabajo de la biblioteca del MIT para poder disponer de una herramienta que les permitiera la implementación de su repositorio institucional. En este sentido se trata de una aplicación que pretende satisfacer las distintas necesidades de difusión, organización y preservación de objetos digitales. En su concepción se tuvieron en cuenta modelos ya existentes, como puede ser el modelo de referencia OAIS^{xvi}, que a pesar de no estar únicamente orientado a la normalización de herramientas tecnológicas, ha tenido un impacto importante en su definición.

DSpace es actualmente el software libre más usado en la implementación de repositorios de software libre^{xvii} y consta con una comunidad de más de 80 desarrolladores conocidos que trabajan en alrededor de 40 proyectos de mejora y ampliación de las funcionalidades de la aplicación. Dentro de la comunidad de programadores, se dispone de un grupo de *Committers* que tiene la responsabilidad última de la actualización del software DSpace, a partir de su propio trabajo o incorporando al código los proyectos que se desarrollan en distintas instituciones. Esto garantiza la continua actualización de la plataforma.

En la actualidad se puede descargar la versión 1.4.1 de la aplicación y ya se está trabajando para disponer de una versión beta de la 1.4.2, que servirá principalmente para incorporar en el código los parches que se han ido realizando de la versión anterior.

Los cambios más importantes se esperan para la versión 1.5, en la que destacará la nueva interfaz de usuario (Manakin^{xviii}) que dotará a la plataforma de una mayor flexibilidad en el diseño de las funcionalidades de la web, ya que el actual modelo resulta, en ocasiones, demasiado rígido. Para la versión 1.5 se espera también la incorporación de un mecanismo que facilite la instalación de aplicaciones desarrolladas para mejorar las funcionalidades de DSpace (AddOn Mechanism), ya que este es uno de los puntos fuertes del programa y del software libre en sí.

A más largo plazo, los puntos de interés en el desarrollo abarcan desde el soporte a la gestión de versiones de documentos, hasta mejorar el soporte multilingüe o incorporar herramientas que faciliten la preservación de los objetos depositados en el sistema.

UPCommons incorpora DSpace en distintas aplicaciones: el propio portal, los repositorios E-prints, Revistas y Trabajos Académicos y actualmente se está trabajando en el desarrollo de un nuevo repositorio de vídeos basado también en DSpace para la gestión de los archivos de vídeo y Flash Video para su publicación.

GUDE: Grupo de Usuarios de DSpace de España

Después de identificar que en España numerosos proyectos de creación de repositorios institucionales tienen como plataforma tecnológica el programa DSpace, se lanzó la iniciativa de crear un Grupo de Usuarios de DSpace de España.

El grupo empezó a funcionar en octubre del año 2006 con los objetivos principales de crear un marco a nivel español para compartir experiencias en la implementación de DSpace, así como participar en su desarrollo. Cuenta en la actualidad con una lista de distribución con aproximadamente 80 subscripciones y un wiki^{xix} que proporcionan soporte a la comunicación de los miembros del grupo.

Entre las principales actividades que se han llevado a cabo hay que remarcar el uso de la lista de distribución para resolver dudas sobre el funcionamiento e instalación de DSpace, ya que muchos de los proyectos se encuentran aún en fase de desarrollo, y la traducción de la interfaz de DSpace a distintas lenguas oficiales de España.

3.2.- Arquitectura LAMP

UPCommons incorpora en algunos de sus repositorios desarrollos propios, como es el caso del Archivo Gráfico de la ETSAB o el UPC OpenCourseWare.

Para estos dos productos se están implementando herramientas de desarrollo propio muy adaptadas a las necesidades de cada proyecto. Se utiliza para su despliegue la llamada arquitectura LAMP abierta:

- Sistema Operativo: Linux
- Servidor web: Apache
- Gestor de bases de datos: MySQL
- Lenguaje de programación: PHP

Los motivos principales para escoger esta arquitectura fueron el amplio uso que se está haciendo en la actualidad de estas herramientas y su clara orientación al desarrollo de servicios web, lo que está favoreciendo también su rápido desarrollo y adaptación a nuevos estándares. Cabe destacar que PHP es un lenguaje multiplataforma, con unos costes de formación para el personal técnico bajo ya que cuenta con una cantidad de funciones y librerías que facilitan en gran medida la tarea del programador. Lo mismo se podría decir del gestor de bases de datos MySQL.

4.- Estándares abiertos

La disponibilidad de estándares abiertos ha significado también una gran ayuda en el desarrollo de los distintos proyectos, ya que nos ha permitido trabajar en entornos normalizados y ha facilitado la interoperabilidad entre distintos sistemas.

De entre los estándares abiertos utilizados hay que resaltar, quizá, tres: Dublin Core, como formato de metadatos para los distintos repositorios; OAI-MHP^{xx}, y SOAP^{xxi}, para facilitar la interoperabilidad entre distintos sistemas.

Todos los repositorios aplican el formato de metadatos Dublin Core^{xxii} para el etiquetado de los objetos digitales que contienen y, si bien existen particularidades en cada uno de ellos, se ha trabajado para conseguir establecer modelos de coherencia entre ellos, normalizando las fuentes para el control de autoridades o aplicando calificadores comunes para cada una de las etiquetas.

Desde sus inicios se ha considerado como factor clave para el desarrollo del proyecto la posibilidad de integrarlo con otros sistemas de información, bien para poder obtener información que facilite el funcionamiento de los distintos repositorios, bien para reaprovechar datos bibliográficos en otros sistemas, como pueden ser el campus digital o el sistema de información de la actividad académica. En este punto ha sido de gran importancia la disponibilidad de estándares abiertos, no sólo para los propios desarrollos efectuados sino también por la facilidad de encontrar herramientas tecnológicas ya desarrolladas en código abierto que han facilitado en gran medida su implementación.

ⁱReferencias:

Consorti de Biblioteques Universitàries de Catalunya. *TDX*.
<http://www.tdx.cbuc.es/>

ⁱⁱ Universitat Politècnica de Catalunya. Servei de Biblioteques i Documentació *UPCommons: portal d'accés obert al coneixement de la UPC*
<https://upcommons.upc.edu/>

ⁱⁱⁱ Universitat Politècnica de Catalunya. Servei de Biblioteques i Documentació. *E-prints UPC*
<https://e-prints.upc.edu/>

^{iv} Creative Commons
<https://e-prints.upc.edu/>

^v Universitat Politècnica de Catalunya. Servei de Biblioteques i Documentació. *Revistes UPC*
<https://e-revistes.upc.edu/>

^{vi}
^{vii} Networked Digital Library of Theses and Dissertations
<http://www.ndltd.org/index.es.html>

^{viii} Universitat Politècnica de Catalunya. Servei de Biblioteques i Documentació *Treballs acadèmics UPC*
<https://upcommons.upc.edu/pfc/>

^{ix} Universitat Politècnica de Catalunya. Biblioteca de l'Escola Tècnica Superior d'Arquitectura de Barcelona. *Arxiu gràfic de l'ETSAB*
<http://eprints.upc.es/adminbustia/bib210/>

^x Universitat Politècnica de Catalunya. Servei de Biblioteques i Documentació *Videoteca digital de la UPC*
<http://biblioteca.upc.es/video/>

^{xi} Pérez, A. ... [et al.]. (2006). Factorías de Recursos Docentes: un servicio colaborativo de apoyo a la docencia en las Bibliotecas de la UPC. Simposio Pluridisciplinar sobre Objetos y Diseño de Aprendizaje Apoyados en Tecnología (3r: 2006: Oviedo). *Actas del III Simposio Pluridisciplinar sobre Objetos y Diseño de Aprendizaje Apoyados en Tecnología*. Oviedo: Universidad de Oviedo, 2006.

^{xii} Massachusetts Institute of Technology. MIT OpenCourseWare
<http://ocw.mit.edu/index.html>

^{xiii}
^{xiv} Borràs, A. ... [et al.]. (2006). Programari de codi lliure per gestionar dipòsits digitals: el procés de tria dut a terme al CBUC. *BiD: textos universitaris de biblioteconomia i documentació*, 16 (juny 2006).
http://www2.ub.edu/bid/consulta_articulos.php?fichero=16borras.htm

^{xv} DSpace Federation. DSpace
<http://www.dspace.org/>

^{xvi} OCLC. (2004). *The Open Archival Information System Reference Model: Introductory Guide*
http://www.dpconline.org/docs/lavoie_OAIS.pdf

^{xvii} Brody, T. Registry of Open Access Repositories (ROAR) : System Software
<http://roar.eprints.org/index.php?action=browse#version>

^{xviii} Manakin Working Group. Manakin
<http://wiki.dspace.org/index.php/Manakin>

^{xix} GUDE: Grupo de Usuarios de DSpace de España
<http://sod.upc.es/gude/index.php/Portada>

^{xx} Open Archives Initiative. The Open Archives Initiative Protocol for Metadata Harvesting
<http://www.openarchives.org/OAI/openarchivesprotocol.html>

^{xxi} W3C. XML Protocol Working Group
<http://www.w3.org/2000/xp/Group/>

^{xxii} Dublin Core Metadata Initiative
<http://es.dublincore.org/>